DOCUMENT RESUME

ED 282 907 TM 870 318

AUTHOR Haladyna, Thomas M.; And Others
TITLE Item Shells: Progress and Potential.
PUB DATE Apr 87
NOTE 20p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, DC, April 21-23, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Difficulty Level; Health Personnel; Item Banks; *Multiple Choice Tests; *Professional Training; Summative Evaluation; *Test Construction; *Test Items
IDENTIFIERS *Item Shells

ABSTRACT

This paper discusses the development and use of "item shells" in constructing multiple-choice tests. An item shell is a "hollow" item that contains the syntactic structure and context of an item without specific content. Item shells are empirically developed from successfully used items selected from an existing item pool. Use of item shells in item writing is a compromise between conventional methods, with considerable freedom in developing items, and the newer approaches, with little or no freedom in developing items. This paper evaluates the item shell technique from the standpoint of training item writers in health care professions. The following criteria were applied in the evaluation: difficulty, productivity, quality, helpfulness, satisfaction, editorial quality, and item performance. Most writers indicated that the quality of their items would remain the same using either method (conventional or item shells). Item shells were found useful for generating items in specific areas of need within a test item pool. Comments about ease of use were positive, though some suggested that the shells be used only in specific situations. The majority of item writers indicated that they could generate more items with item shells. Because the item shells were developed from previously edited items, items that adhered to the shells were editorially sound. The study concludes that further evaluation is needed. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED282907

Item Shells: Progress and Potential

Thomas M. Haladyna
Arizona State University West Campus

Russelyn Roby Shindoll
The American College Testing Program

Lee Langhammer Law
The American College Testing Program

ABSTRACT

Writing multiple-choice test items has been typically characterized as more of an art than a science. However, several recent theoretical developments have resulted in new technologies for generating test items. Conventional item writing has many strengths and weaknesses, as do these newer technologies. The procedure presented, item shells, offers a compromise between conventional methods, in which item writers have considerable freedom in developing items, and the newer approaches, in which item writers have little or no freedom in developing items. An evaluation of the item shell technique is presented from the standpoint of training item writers in the health care professions.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T. Haladyna

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC, April, 1987

BEST COPY AVAILABLE

TM 870 318

Introduction

Writing quality multiple-choice test items is commonly considered a matter of experience, either accompanied by the wisdom passed from mentor to student, or through rigorous study of textbooks on testing (Cronbach, 1970; Ebel, 1951; Nitko, 1982; Roid & Haladyna, 1982; Wesman, 1971; Wood, 1977). An inspection of various sources reveals that information specifying item-writing rules and their validity is inconsistent (Haladyna & Downing, 1987).

Traditional methods of writing multiple-choice items are characterized by the absence of many restraints in terms of actual development. Despite the limitations outlined by over 40 rules for multiple-choice item writing (Haladyna & Downing, 1987), item writers still have considerable latitude in selecting the specific syntactic form of the item.

A number of item-writing technologies have been developed in recent years for the multiple-choice format, including item-writing algorithms (Bormuth, 1971), facet theory (Guttman, 1969), and item forms (Hively, Patterson & Page, 1969). These new item-writing technologies are characterized by a design that enables different item writers to develop similar, if not identical, test items. These new technologies yield a process which is more standardized and uniform than conventional item writing methods, which permit a great deal of creative freedom to the item writer.

A persistent criticism of these newer technologies is a lack of feasibility. Another criticism is that these procedures do not permit enough creative freedom for the item writer. These newer technologies, however, provide the objectivity often thought of as the desirable end to item writing. A compromise between new and traditional approaches to item writing is needed to benefit the science of item writing.

The purpose of this paper is to propose a new technique for developing the multiple-choice test item that represents this compromise between the complete freedom of traditional methods of item writing and the formulaic restrictions of the newer technologies. This tool, called an "item shell," was produced as a result of training health care professionals how to generate high quality multiple-choice test items for their respective certification and licensing examinations. These health care professionals possess expertise in their field, but lack the requisite item-writing skills and relevant experience to produce high quality multiple-choice items with ease. Given that health care professionals have heavy work schedules and various other commitments, the task of writing test items must be streamlined and structured so they are able to produce the best items in the shortest time period. The item shell technique is believed to provide a useful alternative to more conventional item-writing approaches without the strict limitations on creative freedom that the newer item-writing methods appear to have.

The first section of this paper presents the technique, its

justification, features, and examples in a real-life testing program where health care professionals were trained to write test items for a licensing examination program. The second section provides an informal evaluation of the technique from the standpoint of these health care professionals.

Development of the Item Shells

An item shell is a "hollow" item that contains the syntactic structure and context of an item without specific content. Item shells are empirically developed from successfully used items selected from an existing test item pool. The item shell provides the item writer with a truly infinite number of possible items while regulating the item-writing process. With the syntactic structure of an item complete, the item writer may concentrate primarily on what specific, important information should be tested rather than grappling with whether an item should begin with "Which of the following," "What," or "How." Providing a basic structure also supplies the item writer with a springboard for other possible structures or variations of structures. Consequently, the item shells help item writers create original items as well as variations of a single item. However, item shells should serve as guides to generate ideas, and should not be presented as the only acceptable way to write items. The danger in the latter is obvious, as most item writers object to the perceived limitation in creative freedom.

The item shell technique also enables an item writer to

select the taxonomic level to be tested. While the use of the "hollow" item, which is stripped of only specific content, is clearly suited for testing at the recall or comprehension levels, an investigation of a more open structure indicates that shells may be used to test higher-level thinking. The structures provided for lower-level thinking are complete questions or open-ended statements with only a condition or term missing. However, the structures for the higher-level concepts are modified to allow a greater number of possibilities and increased creative freedom for the item writer. The item writer is given a basic structure and context along with several "prompts," which suggest a variety of possible routes of development for an item. While these open structures require more than filling in the blanks, they provide direction for item writers who claim writer's block or for those who feel comfortable writing on a limited number of subjects within their content expertise.

The item shells were originally developed to help item writers who found writing high quality, multiple-choice items difficult. A variety of successfully used items, from both lower and higher taxonomic levels of cognitive behavior were then selected from an existing test item pool and stripped of specific content. From items at the lower taxonomic levels, only a single word or phrase and the options were deleted, as shown in

Figures 1 and 2 (Note: all sample items included arehypothetical).

The abbreviation MRI refers to which of the following?	The abbreviation _____ refers to which of the following?
A. Magnetic reaction imagery	A.
B. Magnetic resonance imaging	B.
C. Marginal resonance imagery	C.
D. Marginal refractory image	D.

Figure 1 Figure 2

After specific content is removed, the syntactic structure remaining, shown in Figure 2, frees the item writer to concentrate on the material, or any abbreviation as in this case, rather than on how to word the stem. Item shells may also be used to help generate independent items by associating words or concepts in the shell with ideas for other items, as shown in Figures 3 and 4.

The abbreviations(____ ,____ , and _____) or (____and____) are most likely to appear on:

A.
B.
C.
D.

Figure 3

The appropriate abbreviation for _____is:

A.
B.
C.
D.

Figure 4

When the original item (Figure 1) is stripped of specific content, any conditions, information, or variations may be inserted to create any number of new items. Figures 1-4 may show simple abbreviation items; however, the possible item shown in Figure 3 requires that the examinee know the appropriate location (e.g., medication record, reference book, in virgile construction) for specific abbreviations. The possible item shown in Figure 4 is simply the original item written in reverse.

Item shells developed for higher-level thinking tasks do not have the rigid structure of the shells shown in Figures 1-4. The open shell provides direction and, through the use of "prompts," suggests several possible avenues of development. When the item shown in Figure 5 is stripped of most of its specific information, it is left with direction and "prompts," as shown in Figure 6.

A 6-year-old child is brought to the hospital with contusions over the abdomen and chest as a result of an automobile accident. Initial treatment should consist of:

A.
B.
C.
D.

Figure 5

A (age, pre-existing condition, sex of patient) is brought to the hospital (complaining of, with injuries showing symptoms of) (as a result of). Initial (treatment, intervention, diagnosis, laboratory studies,

antibiotic therapy) (should consist of, is, are, include):

A.
B.
C.
D.

Figure 6

The skeletal structure of the open item shell offers the item writer a variety of possibilities for development based on the premise "A patient is brought to the hospital with a specific condition(s) which requires some sort of attention or action." The premise is simple, yet allows the item writer the opportunity to develop as difficult and complicated or as simple an item as desired.

Use of the Item Shell

Item shells can be used in a variety of situations and for any number of specific purposes. The shells used in this study were first offered to item writers at a time when the existing test item pool lacked items in specific content areas. The performance of the existing items in these content areas was evaluated before the item was considered a shell candidate. Items that performed successfully were then stripped of specific content or modified to become open shells and placed in a single document, along with a statement outlining the specific content to be tested. No more than ten shells were offered in each content area to avoid overwhelming the item writer. Again, the

primary purpose of the shells is to guide, not complicate, the process for the item writer.

After all of the items were carefully modified to form shells and placed with a content statement, the shells were formatted into a resource document for item writers. The shells were distributed with blank test item writing forms to item writers who, for the most part, had rough draft forms of items. The shells, then, may have another potential use--as a polishing tool. While many item writers have rough draft forms in mind or on paper, final polishing of items is often problematic. Because the shell provides a polished syntactic structure, the item writer may simply incorporate the rough draft information into the shell or use the shell as a guide for polishing the original rough draft. Or, if an item writer knows that items about specific conditions (e.g., AIDS, Parkinson's disease, etc.) are needed in the test item pool, the open-structured shells may provide a context to establish such an item.

Item shells may also complement item-writing instruction. It is not often possible to assemble groups of inexperienced item writers to teach the principles of item writing; therefore, well-constructed item shells accompanied by sample items may be useful tools in the absence of guided instruction. Also, item shells and examples may be appropriate ancillary materials for follow-up after instruction. The item writer, then, would not have to rely upon memory or sketchy notes when generating items six months after instruction.

The item shell approach to item writing complements the more traditional methods of item writing without restricting item writers to any single form. To briefly reiterate, item shells provide syntactic or contextual alternatives that traditional item-writing methods do not. The task for the item writer, then, shifts from primary construction to specification of content for certain items or "fleshing out" more complex items. Because the item writer, ideally, toils less with expression, item writing may be a faster, more efficient process and less likely to be regarded with the enthusiasm afforded the dentist's drill. For many individuals, writing may be associated with punishment, either as "I will not talk in class" one thousand times or an essay on the essence of the "Pledge of Allegiance." Item shells may be used as the proverbial "if I could just get a start" to item writing.

An Evaluation of Item Shells in a Test Development Setting

The evaluation study was intended to examine item shells in a "real life" item development setting. The criteria for evaluating item shells is given below.

1. Difficulty, which was the extent that item writers perceived that item shells made item writing easier;

2. Productivity, which was the extent that item writers perceived using item shells yielded more items than if this item-writing aid was not used;

3. Quality, which was the extent that item writers perceived using item shells yielded items of higher quality than if item shells were not used;

4. Helpfulness, which was the extent that item writers perceived that item shells helped in the item-writing process;

5. Satisfaction, which was an overall recommendation for or against item shells;

6. Editorial quality, which was an assessment by professional test editors regarding the quality of items vis a via concerns such as clarity of expression, spelling, punctuation, grammar, and adherence to basic item-writing rules; and

7. Item performance, which is the tendency for items to perform as they were intended (generally item difficulty in discrimination fall within certain limits).

****
Insert Table 1 here
****

The form used to elicit item writers responses to the first five responses is presented in Table 1. Editorial quality was evaluated through an analysis of items prepared using item shells and items prepared in the conventional manner; these judgments came from other staff members who professionally edit these items. While admittedly, this is subjective criteria, it is taken as possible convenient evidence to increase the validity of any conclusion about the value of using item shells.

Quality

While all item writers believed they could produce items as quickly with either method, their views about the quality of the items generated using the shells differed significantly. Most item writers indicated that the quality of their items would remain the same using either method. Some also felt the quality of the items they produce would improve only through the use of conventional methods, One item writer expressed concern about the possibility of generating too many "clone" items, and feared that the extensive use of the shells would undermine the creativity and individuality of item writers and ultimately the examinations themselves.

The perceived danger in using item shells only is that after several cycles of item development, a single 200-item test could be comprised of variations of a few item structures. This concern about "cloned" items reinforces a basic premise for using the item shell as a tool: item shells should serve only as guides to generate ideas or provide direction for development. Since each shell is based on successfully performing items, it is thought that items derived from these shells might also perform successfully. The ultimate determination of the issue of quality is an empirical one.

Helpfulness and Satisfaction

The item writers believed the item shells to be particularly helpful for generating items in specific areas of need within a test item pool. For example, an item writer who is assigned 12

Results and Discussions

Difficulty

Overall, comments about ease of use were positive, though some item writers qualified their evaluations by indicating that the shells be used only in specific situations. The health care professionals who evaluated the item shells were all experienced item writers. Most indicated that they had little difficulty writing items using conventional methods, but recommended that the item shells be given to inexperienced item writers or introduced for use with final rough draft forms of items to reduce the need for editing.

Because the health care professionals who evaluated the shells were experienced item writers, they may have viewed the introduction of the item shells differently then first-time item writers. If the shells had been evaluated by first-time item writers, perhaps the perception of difficulty would have been less conservative.

Productivity

Comments about whether item shells help increase productivity were emphatically positive. The majority of item writers indicated that they could generate more items with the item shells, particularly when faced with writing items in areas that lack enough items or when "stuck" generating items in areas other than their specific expertise. One item writer commented that item shells would most likely increase productivity and quality for less proficient item writers.

items about drug classifications and 10 items about treatment of overdosage would be able to refer to the shells specific to these two content areas for items or for ideas for items. When given a specific content structure, most of the item writers indicated that the shells were helpful for generating ideas and would recommend that the shells be used to complement the traditional methods of item writing.

Editorial Quality

Of the 100 items reviewed, approximately 50 were written in the content areas for which item shells were provided. The item writers, in the course of routine item development, were asked to rebuild specific areas of the test item pool, areas that lacked a significant number of high quality items. Only approximately 20 of the 50 items in these areas were written using the item shells. The other thirty were noticeably different or classified incorrectly.

Because the item shells were developed from previously edited items, new items that strictly adhered to the shells were editorially sound. More specifically, the items were grammatically correct even when other items written by the same individual were grammatically incorrect. Spelling was not affected to a great degree, because item writers traditionally misspell words that would be considered specific content, such as diseases, drug names, or methods of treatment. The shells provide only simple words, such as "Which of the following," "patient," etc. Punctuation, especially of open-ended items, is

commonly omitted by most item writers. But in the items generated with the shells, item writers remembered to insert terminal punctuation for both the stem and the options more often. However, in items generated without the shells, terminal punctuation appears with less frequency.

Item writers generated items using the shells for the lower taxonomic levels. None of the item writers attempted items using the open shells, suggesting the need for more refining or more instruction on how to use the open item shells. It is important to note that the overall performance of item writers, that is, all items generated, did not improve even with item shells as a guide for sound item-writing practice. So, the item shells should not be considered a means to improve overall performance, although they may be used in this capacity in the future.

Item Performance

Item performance data is not available as yet, but plans have been made to trace the history of items produced using item shells in an effort to determine the merits of these items in long-term development, which includes pretesting and actual use in a professional licensing examination program.

Conclusions

A new technique, item shells, has been introduced and described as a procedure which augments conventional item writing. Essentially, the technique is empirically based, because high quality items are taken from a test item pool and stripped of specific content to form the shell of the original item. It is presumed that the essential syntactic structure of a high quality item is useful for generating items with similar syntax which retain the good qualities of the original item.

The usefulness of item shells needs further evaluation, particularly with respect to item performance and with other groups of item writers, perhaps outside the health professions. The item shell appears to have applicability to a variety of subject matters and should be a useful complement to conventional item writing, and particularly for inexperienced item writers.

REFERENCES

Bormuth, J. R. (1970). On the theory of achievement of test items. Chicago, IL: Univ. of Chicago Press.

Cronbach, L. J. (1970). Review of "On the theory of achievement test items by J. R. Bormuth. Psychometrika, 35, 509-511.

Ebel, R. (1951). Writing the test item. In E. F. Lindquist (Ed.) Educational Measurement, (1st ed.). Washington, DC: American Council on Education, 185-249.

Guttman, L. (1969). Integration of test design and analysis. In Proceedings of the 1969 invitational conference on testing problems. Princeton, NJ: Educational Testing Service.

Haladyna, T. M. & Downing, S. M. (1987). The validity of a taxonomy of multiple-choice item writing rules. (Submitted for publication)

Hively, W., Patterson, H. L., & Page, S. A. (1968), A "universe-defined" system of arithmetic achievement tests. Journal of Educational Measurement, 5, 275-290.

Nitko, A. J. (1984). Book review of Roid and Haladyna's A Technology for test-item writing. Journal of Educational Measurement, 21, 201-204.

Roid, G. H. & Haladyna, T. M. (1982) A technology for test-item writing. New York, NY: Academic Press.

Wesman, A. G. (1971). Writing the test item. In R. L. Thorndike (ed.) Educational Measurement. Washington, DC: American Council on Education.

Wood, R. (1977). Multiple choice: A state of the art report. Evaluation in Education: International Progress, 1.

Table 1

Evaluation

☐ Item writing is easier for me if I use shells than if I use conventional methods of item writing.	☐ I can write items just as easily either way.	☐ Item writing is easier for me if I use conventional methods of item writing than if I use item shells.
☐ I can generate items more quickly if I use item shells than if I use conventional methods of item writing.	☐ I can generate items just as quickly either way.	☐ I can generate items more quickly if I use conventional methods of item writing than if I use item shells.
☐ Generally, the quality of my items is improved with the use of item shells for construction.	☐ The quality of my items is the same either way.	☐ Generally, the quality of my items is improved with the use of conventional methods for item writing.
☐ I recommend the use of item shells for item writing.	☐ I recommend both the use of item shells and the use of conventional methods for item writing.	☐ I recommend the use of conventional methods for item writing.

	Very Helpful		Helpful		Not Very Helpful
I found item shells to be	1	2	3	4	5

Briefly describe your experience in using the item shells for generation of test items for the NABPLEX examination.